Nicasio Catanglan Jr
THINKFUL STUDENT

PROPOSED FINAL CAPSTONE

PROJECT NAME:
 U.S. ELECTRIC POWER PREDICTION FOR THE NEXT 20 YEARS USING TIME SERIES

OVERVIEW:
Future patterns of changing global energy system at a time of increasing uncertainties and finds that major transformations are underway for the global energy sector, from growing electrification to the expansion of renewables, upheavals in oil production and globalization of natural gas markets. Across all regions and fuels, policy choices made by governments will determine the shape of the energy system in the future.

GOAL:
The goal of this project is to build a Time Series model that can predict the total estimated value of Revenue in Dollars, Power Demand in MWH, Number of Customers and Cost per KWH for U.S in the next 20 years.

HOW DOES THIS PROJECT HELP?
Power Plants, Electrical Companies, Government Agencies and Local Power Producers can utilize this model to predict and monitor power demand and supply in real time to be able to lay out plans in advance to meet the said Electrical Power supply and demand. This will also help to foresee the said increasing changing global energy system uncertainties.

ABOUT THE DATA:
I got the data from the US Energy Information Administration. The URL to access the data "https://www.eia.gov/electricity/data.php" The data is about the monthly revenue, energy consumption in mwh, Number of customers and price per kwh from 1990 to present. Each data for Residential, Commercial and Industrial is collected separately from all the US State and join them together in one data reference to time intervals. This is one of the challenges in joining all the data and put them together in accordance to the date it was collected. Cleaning the data is one I expected to be a challenge. In this final capstone, I will apply most in what I learned from the program like python, statistic, analyzation, cleaning of data, visualization, testing, feature importance, model application particularly ARIMA for prediction.

METHODOLOGY/APPROACH:
1. Import needed Libraries like Panda, Numpy, Matplotlib, Seaborn, Sklearn toolkit etc. Data is downloaded to DataFrame for manipulation using pandas. Data wrangling will be performed on the dataset. Data wrangling involves renaming columns, joining multiple data into one according to datetime, cleaning null values and filling it with appropriate methods (i.e., value replacement, drop columns, drop rows etc.).
2. Visualization and Data Analyzation to know more about the data. Using Statistics and visualization toolkits from Matplotlib and Seaborn, to be able to know the outliers, trend, seasonality, stationarity, significant feature and important order value for ARIMA of the Time Series data.
3. Finally, testing ARIMA orders and choosing the best model then we predict and evaluate the performance of the model.